The f-widening Theorem
Cantor-Dedekind-Zermelo

R

Introduction

This is the twin article of the one titled “Impossible Machine Theorems”.
That presented two theorems never mentioned in text books and here again we will state one that is unnamed. If they were not relating to the most important theorems of twentieth century mathematics we shouldn’t care that much.
But just as Rice Theorem is the most important conceptual realization in the field of Effectivity, the Well-Ordering Theorem was the most important one in Set Theory.
So if something unnamed lurks behind these then we must be suspicious.

Cantor’s two proofs that the fractions are sequencable while the infinite decimals are not, is the entrance for anybody into Set Theory. But the jungle that this so simple idea of the sets creates is not revealed by these two amazing arguments at all.
So by now, I think that a formal approach should be starting with the following three:
$\{a,b\} = \{b,a\}$, $\{a,a\} = \{a\}$, $(a,b) = (c,d) \leftrightarrow (a = c \ , \ b = d)$.
Surprisingly, the crucial $\in$ membership relation is not even appearing.
But these make it clear that the { } collection idea is ignoring all structures of the members that even our writing of the expressions can not do.
We also introduced the round bracket for ordered pairs that at least reflects this minimal structure of an order. Of course, the whole miracle of sets is that in spite of its sturucturelessness it can describe all structures. And thus, already here at this minimal level we should show how different ideas can do this for the ordered pair.
That is, we should show the different possible ordered pair replacements.
(a,b)-short = $\{a,\{a,b\}\}$, (a,b)-Kuratowski = $\{\{a\},\{a,b\}\}$, . . .
The dots mean that there are even further possibilities and thus shows that we entered a labyrinth of ugly formal solutions to beautiful ideas.
The belief that the beautiful ideas are a reality or not is then up to anybody himself.
For a Platonist like me it is so. In fact, now I also believe that the uglinesses of the formalizations is also a reflection of an even deeper second abstract reality.
Cantor was not aware of this whole line of thoughts that I just told.
After he realized what God handed to him the single membership relation that can describe everything, he introduced equivalence $\sim$ the relation to compare infinities.
In fact, that was his main goal. Of course the objective truth that he started to chew on something much bigger than he was aware of, came through. Namely, by that he became the father of Topology too. So he investigated the continuous sets too.
He discovered the Cantor sets and even arrived at some point sets where the two directions the infinity comparability and the topological features meet.
But this was a meager success actually showing even more that the two directions are not in accord. Even more amazingly, there was a third line, a side line of the comparability of sets that hid the biggest reason why we have a fundamental problem.
This was tackled only by Zermelo though Dedekind the earliest supporter of Cantor made a discovery that related to what Zermelo achieved, the Well-ordering Theorem.
So the three names under our title are explained. The essence of the article is why the role of Dedekind is not realized by anybody and was not even by Zermelo himself.
And the simple explanation for this ignorance is that the above mentioned success of Zermelo was not merely a smart new proof but the realization of a new axiom as well.
Unfortunately, the two big things got so intermingled that a third got buried.
The depth of that will not be explained in this article and yet the article is complete.
Just as the twin article made a crystallization of the post Rice level of effectivity, this article will make a crystallization of the Well-ordering Theorem.

All this what I just foretold shows that to achieve a didactical correctness is very hard. Historical exposition is definitely not didactical and so now I will make a big jump in time and start from the point when it indeed became proven that Set Theory is the theory of everything. The name of the theorem became the Completeness Theorem.
In 1930 when Gödel presented his proof for this completeness of mathematical logic Hilbert the father of this logic was present but he probably couldn't follow the proof. But in the intermissions of the congress Gödel was telling about something even more difficult to comprehend to the understanding ears, that practically meant Neumann.
That difficult concept later became The Incompleteness Theorem.
So in that 1930 math congress both the completeness of our logic and the incompleteness of the axiom system for Number Theory were "floating in the air".
The first as a culmination and the second as a seed.
By what I just said, it is obvious that these two, The Completeness Theorem and the Incompleteness Theorem are not contradicting each other because the first is about Logic while the second is about Axiom Systems.

The Root

The mentioning of Number Theory above would make us think that a common root of the two results and thus a clarification of the two meanings of the words "complete" one for Logic and one for axiom systems could lie in the formalization of Number Theory by Peano. But the real root lies in Geometry.
Just as our life is determined by subconscious mysterious factors, the awakening of the scientific abstract concepts are also far from logical continuations.
So Cantor would be very shocked by what I will reveal now, what "initiated" his most important abstraction of the human mind. Even more surprisingly, an other just as hidden trail was present behind the mentioned axiomatization of Number Theory.
That went back to Aristotle's Formal logic, namely his recognition of the presently called two quantors in mathematical logic. These two parallel lines had no relation for two thousand years and only became connected finally in New Math.
What I call as the "root" in the section title, is a recognition of Beltrami that clearly necessitates not only the existence of a Logic but Sets relating to such Logic.

Points, lines and circles were the same basic concepts of Geometry as they are today in elementary schools. But today when the concept of sets slowly entered elementary education, an also present algebraization is pushing out Geometry altogether.
This is tragic because sets have a seemingly very natural place in Geometry.
The lines and circles are sets of points and their intersections are simply common members of them as sets. This is so evident that there is only one thing more illogical than pushing out Geometry from the education system. Namely, why the Greek mathematicians didn't recognize the concept of sets already!
But as I said, there is a subconscious logic in the history of concepts and it seems this includes subconscious denials too!
Parallelity is the crucial concept of no intersection of two lines in a plane.
But here the non common member of them as sets is the least important as clarification. Instead, two alternative conditions can be observed.
One uses the concept of distance, the other of the angles.
Indeed, parallels have the same distance between them which implies trivially the non crossing and also they must have the same angle to any third line that crosses them.
This second does not imply a non crossing instantly but this is just the start of the complications. How the three conditions, the non crossing, the fix distance and the same directionality relate is a mess. There is a hidden logical simplicity though.

Namely, just regarding a single line and a point outside, the assumption that there is only a single second line in the plane of the first line and the point that doesn't cross the first, implies that the mentioned three conditions all imply each other.
So we would assume that Euclid used this singular non crossing assumption as an axiom of parallelity. But he didn't! Another subconscious denial?
He used a fairly complicated claim instead, that he himself was not too happy with.
Since this was the only ugly one among the geometrical axioms, mathematicians tried to avoid it by finding some tricky proof of it from the simpler beautiful axioms.
The most successful method to find a "hidden" proof is "indirectness".
We assume the opposite of what we claim and try to derive a contradiction.
Then usually we can transform the chain into a direct proof of the original claim.
With the modernized axiom of parallelity finalized by Playfair around the end of the 18-th century, the indirectness means that we imagine more non crossing lines with a fixed one through an also fixed point in their plane and try to find a contradiction.
This is exactly what Bolyai did but instead of a contradiction wonderful possibilities opened. He wrote: "from nothing I created a new world".
He sent his results to Gauss who was a fellow student of Bolyai's father.
Gauss' rude reply devastated Bolyai but as rude as it was, it was true.
He indeed kept in his drawer the same earlier recognitions.
Both Gauss and Bolyai thought that a world where more non crossing lines are possible has some kind of physicality. Bolyai even tried to invent some method by which light rays as lines could confirm this non uniqueness.
But why should light rays be the lines? And what if light rays are bent for some other reasons? So venturing into Physics is a dangerous step.
But most importantly, we don't have to! The wonderful new world is mathematical.
A step towards this recognition could have been if they had made a simple abstraction about a "new world" that was already very well known by both Gauss and Bolyai.
This was the so called spherical "plane" geometry on the surface of a sphere.
The points with fixed distance from a fixed point are obviously circles here too.
Of course, these circles have maximal possible size that are the "main circles".
Like the equator or the time zone "lines" on our globe.
The big recognition is then that these main circles could be regarded as abstract lines.
But obviously the basic axiom that two lines cross in a single point would be false.
Well, let's make it true! All we have to do is make the diagonally opposite points as single "points" of our new geometry. So Madrid in Spain and Weber in New Zealand would be regarded as a single point and similarly every point would include its diametrical opposite. And then in this strange "plane" the unique non crossing will not be true! Simply because there are no non crossing lines at all! Indeed, every two main circle cross in such diametrical two or now rather "single" point.
Unfortunately, since we kept the concept of distance almost the same, we will still not have all the basic axioms to be true in this world. Indeed, there are axioms that express the infinite length of the lines. This also suggest though a solution by distorting the distances so that in our "model" an infinity of the lines remains too.
When this most drastic trick of the distance distortion was recognized by the German mathematicians to visualize models in a disc for example so that the strings of the disc could be regarded as infinite lines, none of them realized that in less than a century physicists will say that the universe is finite and yet from the inside it is infinite.
The physical realization of this contradiction is an expanding universe with relativistic distances. But the essence is the same. Just as going towards the edge of a disc that models an infinite plane we feel from the inside that the strings are infinite, similarly we can not go to the edge of our finite universe. So after all, Bolyai was not so crazy to investigate light rays as lines but it was a too early, half baked idea yet. Most importantly it jumped over the immediate and more important recognition of models.

But actually those smart German mathematicians that envisioned these non Euclidian models of Geometry still missed something very important and the coin only dropped for the Italian Beltrami. Namely, that these models prove something original too.
What was the original goal? To try to find a proof for the parallelity axiom.
The indirectness showed wonderful new worlds where the axiom is false but the new models are worlds that “prove” that such proof of the parallelity axiom is impossible!
Indeed, if there were a proof then it were a logical necessity and so in all models where the basic axioms are true the parallelity axiom would have to be too.
But they were able to create some models where all the basic axioms are true yet either there is no non crossing line or there are more. The hidden assumption of this argument is of course about the still not clarified concept of Logic and Model.
And yet their relation was observable! That’s how Beltrami necessitated Logic and Sets. In fact, he necessitated what a completeness of Logic should mean too!
His argument only assumed that models must obey logic but this raises the opposite question, what the meaning of a logic obeying models could be.
And the answer is pretty straightforward. Namely, that if from some axioms a claim does not follow by logical derivation then there must exist some model where the axioms are true but the claim is false. So for a set of axioms and a non derivable statement we must be able to construct a model that exemplifies this.
This is what the completeness of a Logic means and this is what Gödel demonstrated in 1930 about the logic developed by Hilbert.
And he used the concepts of Zermelo which this article will explain.

Comparing infinites

But we must return to Cantor who had no idea about how his grand abstraction will also solve the exactification of models. Indeed, his original goal, was “merely” the comparing of infinities. Sadly, his goal in its fullness failed.
We can not compare the wild possibilities of infinites that Cantor demonstrated.
But neither can we say that he failed. The partial success is what we’ll deal with first.
The reasons of the bigger failure are yet hidden to us. Some believe that it is meant to be like this, some believe that there comes a perfect description of infinities.
One thing is sure! That this partial success will be part of any later clarification.
So all that follow is very important.

That the $\sim$ equivalence relation is something heuristic, can be best understood by the following experiment. We have a ball room filled with boys and girls.
Some dance, some just wait to be asked, some shy some throw flirty glances.
The teacher wants to know if the boys or the girls are more in number.
Counting both would be a solution. Long and hard to execute. But instead she can just shout: Let’s form pairs! And then, since it’s an order not a choice, it can happen.
Of course it’s very unlikely that no boy or girl would remain without pair.
And whichever is the case will tell at once who were more in number.
So the idea of equivalence is useful already at finite sets.
The shocking new feature that comes with infinite many boys and girls is that a pair forming equivalence that exhausts all boys or girls will not tell that the other not exhausted set is more. Simply, because we may have an other pair forming equivalence where the exact opposite is the situation or where nobody stays unpaired at all. And to see this is easy if we regard the odd natural numbers as boys while the evens as the girls.
Probably this level of consciousness was the world about infinites before Cantor.
So it was an unasked question if there are different infinities.
Or it was left to the appearance of the structured sets to make this judgment.

The fractions obviously contain all whole numbers as special fractions.
But even just the proper fractions under 1 can trivially contain all n naturals if we regard the $\frac{1}{n}$ reciprocals with $n > 1$. All the other non 1 numerator fractions remain unpaired in this assignment or pairing. And they are so much more!
So are the fractions more than the naturals? As Cantor showed, no! Look at this:

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \ldots$$

Embarrassingly simple! We listed all proper fractions by simply going through all increasing denominators and list the finite many possible numerators.
But we can do something much better, with same method, namely sequence all fractions as:

$$\frac{1}{1}, \quad \frac{1}{2}, \frac{2}{1}, \quad \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \quad \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \quad \frac{1}{5}, \frac{2}{4}, \frac{3}{3}, \frac{4}{2}, \frac{5}{1}, \ldots$$

Now the finite groups are by the increasing totals of the numerator and denominator.
Amazingly, the same trick can be used for something even more drastic.
It is actually the first application of Set Theory to an earlier classical math problem.
Namely, to prove that there are non algebraic or so called transcendental numbers.
In fact, this argument shows also that this transcendental name is not justified at all because the non algebraic numbers are the majority. An algebraic number is one that is root of a polynomial equation with whole coefficients. To sequence all algebraic numbers is a bit more complicated than the fractions above but can be done by the same method, that is to form longer and longer finite groups.
With one T total of the order and the used absolute values of the coefficients, we only have finite many possible polynomials. Plus every n-th order polynomial can have only maximum n many roots. So the possible roots are finite for a T too.
Thus we can list all algebraic numbers again as finite groups with increasing T.
Of course, to see that thus we must have non algebraic numbers, we must enter the second blast that Cantor showed. Namely, that we can not list similarly all the decimals under 1. Indeed, suppose we had them now under each other as:

. 3 4 1 0 7 9 0 1 6 . . .

. 0 9 1 8 5 3 2 4 9 . . .

. 3 6 0 3 7 4 1 1 2 . . .

.

Let's make a D diagonal decimal from the diagonal members:
D = . 3 9 0 . . .
We used the first digit from the first in our list the second from the second, and so on.
It would be very strange if this D were somewhere in our list but nothing forbids it.
But now let's add 1 to all digits of D with using 0 as 9+1. Then this is:
D^{+1} = . 4 0 1 . . .
We can be sure that this D^{+1} anti diagonal number is not in our list. Indeed:
It can not be our first since it has a different first digit.
It can not be our second since it has a different second digit. And so on.

Now come some "depthifications"! Let's see first the fractions.
We obtained all fractions no problem with that, in a nice simple row like the ducks.
But we obtained actually so much more! Indeed, all fraction values were repeated infinitely. Why this redundancy? Well, it is not accidental! A sign of something much deeper, lying in the field of Effectivity. So is Effectivity involved?

No! We can just say, ignore the repeating fraction values and then we listed merely the fraction values. Then of course our list would be much more complicated to tell.
Now the decimals. First of all, can our merely shy remark that D being in the list would be very surprising, somehow be exactified? Can even some chance value to this be assigned? Very heavy stuff! And yes it can be approached but very hard.
This has an amazing sidetracking.
A very old and important question was how the fractions and the decimals relate as sets. And this is a very provocative way to say it because the Greeks had no idea about the infinite decimals. This is a silver platter following from the decimal system. But we didn't lie because the Greeks did raise the question of how the distances relate to the fractions. And today we see that the distances are all infinite decimals.
The length of a table is measured by the cabinet maker in centimeter with knowing that it can be refined by millimeters then tenth of millimeters and so on.
The exact length is an infinite decimal.
The crucial step in solving what the Greeks were just walking around is the division algorithm, taught in all elementary schools. We bring down digits to the remainders:

$$\frac{25}{14} = 25 : 14 = 1.7\underbrace{857142}_{\text{repeating period}}85\ldots$$

```
110
 120
   80
   100
     20
      60
        40
        120
```

The essence is that since the remainders can only be finite kind, we must have a repeating period. And so a fraction can only be such periodic decimal.
So obviously there are non fractional decimals because there must be non periodical decimals. And here comes in the similar observation. If we take a random decimal it wouldn't be periodical for sure. But what is a random one?
Can we get out of this depthification at least? Yes we can. Because if you only care about if there is non periodic decimal you can manufacture one too.
Observe also that Cantor's two proofs are actually a new way to show that some decimals are not fractional. Indeed, the fractional ones are only a sequencable subset.
By the way, the fact that the full set of decimals is non sequencable and the rationals are sequencable, does prove the existence of irrationals but it doesn't prove that their difference set, that is the irrationals is non sequencable too. This relies on the almost trivial step, that two sequences together would be again a sequence.
Indeed, we can go through alternatingly. Then an easy argument can also show that finite many sequences together are sequencable too.
But amazingly, the sequencability of the fractions actually hid the method that shows that a sequence of sequences together is also sequencable:

$$\frac{1}{1},\ \frac{2}{1},\ \frac{3}{1},\ \frac{4}{1},\ \ldots\ldots$$

$$\frac{1}{2},\ \frac{2}{2},\ \frac{3}{2},\ \frac{4}{2},\ \ldots\ldots$$

$$\frac{1}{3},\ \frac{2}{3},\ \frac{3}{3},\ \frac{4}{3},\ \ldots\ldots$$

We can not leave the depthifications yet because a nuance will be our business now.
First of all, our defied sequencing of the decimals raises the question why this was about an equivalence like pairing the boys and girls.
After all, we didn't require such uniqueness for our listed decimals.
At the fractions we said that the repetitions, the redundancy is a side effect of something later in Effectivity. But at the decimals we claimed the negative.
So a repetition among the listed decimals is actually irrelevant. If it is impossible with possible repetitions then it is also impossible without. So it's more about how the equivalence as crucial concept is true or not. And this is right at the first mentioned nitty gritty of the pairings. Indeed, if we pair uniquely members of two sets A and B that have no common members then simple {a,b} pairs are sufficient with the requirement that for an a there can only one b and similarly in reverse. Just like the boys and girls as pairs didn't have to be ordered which gender is the first or second.
So such "raw equivalence" is a more elemental concept than the ordered pair.
But in our listings it was not like this. The naturals dictating this order are the first members and the listed members are the second. Such set of ordered pairs is a function. In a sense the most important concept that we didn't even mention yet.
So our list of some decimals above was actually a function defined on the naturals and taking up values among the decimals.
And what we showed was that the range of the function, that is the set of values taken up, that is the set of second members of the ordered pairs, can not be all the decimals.
This is a much nicer result than our original goal that the set of decimals can not be equivalent to the naturals. No indirectness, no negativity, just the simple fact that for any sequence of decimals, there must be some decimals not listed.
But is this a general reduction of non equivalence to merely functions? Yes, but!
The start is that there are two alternative intuitive reasoning for $A \leq B$.
Which simply means that $A < B$ or $A \sim B$.
The first is saying that A can be "injected" into B by an $\sim$ equivalence.
Which means that there is some B' that $A \sim B' \subseteq B$.
Where $\subseteq$ is the symbol for B' being in B and simply means that $b \in B' \rightarrow b \in B$.
The second is saying that B can "cover" A by an f function.
Which means that $f[B] \supseteq A$.
Here $f[B]$ means not what f assigns to B rather the set of all f(b) assigned values for the b members of B, that is: $\{ f(b) ;\ b \in B \}$.

The wonderful fact is that the two descriptions coincide.
Firstly, A being injected into B by $\sim$ trivially implies the second description by simply regarding as f the $\sim$ equivalence as a function from B' to A and then just extend this by assigning arbitrary members of A to the B – B' members of B.
The reverse is more difficult! From the f function by which B covers A we must make an equivalence. We must regard every $a \in A$ and pick a b* member from those b-s at which f(b) = a. This then obviously becomes an equivalence and still covers A and so the reversal of f becomes an injection of A into B.
But to make these simultaneous picks, one actually needs the Axiom Of Choice.

Now a third intuitive definition can be that $A < B$ should mean $A \leq B$ with the added condition that $A \sim B$ is impossible.
At the naturals as A versus the decimals as B we indeed proved that $A \sim B$ is impossible but strangely we actually showed instead that $B \leq A$ is impossible by showing that the naturals can not cover the decimals by any f.
This of course was enough because it implies the impossibility of a perfect equivalence too. Also strangely, we didn't bother to show $A \leq B$ because it was so obvious by for example the injection to order for any n the .n decimal.
But what is the situation in general? Surprisingly, it "could be" much simpler!

The "could be" means that we need something very expectable first to be proven. Namely, that at least one of $A \le B$ or $B \le A$ always stands for any two A , B sets. We could call this the weak comparability of sets. And indeed, if this is proven then quite simply, an impossibility of $B \le A$ at once implies that $A < B$. So our injection of the naturals into the decimals was not necessary either. The impossibility of $B \le A$ automatically implies the impossibility of $A \sim B$ and also that $A \le B$ must be true.
So the weak comparability of sets is crucial both intuitively and practically.
And our whole article is actually the proof of this weak comparability of sets.
In fact, I will show again how beautiful the new simpler definition of $A < B$ becomes with using the weak comparability. As we just said it is the impossibility of $B \le A$.
But what does that mean? That for every f function $f[A] \supseteq B$ is impossible.
But what does that mean? That for every f function there is some $b \in B$ not in $f[A]$.
Which can be said even more visually as $B - f[A]$ being not empty.

By the way, there is a second expectation hinted by our "at least" in the weak comparability. Namely, that $A \le B$ that is $A \sim B' \subseteq B$ and $B \le A$ that is $A \sim B' \subseteq B$ should imply $A \sim B$. Luckily this was shown quite explicitly by Bernstein.

A naïve idea for the wanted weak comparability of an A and B sets could go like this: We pick an arbitrary a member from A and similarly a b from B.
We assign these to each other as the start of our planned equivalence.
We remove these members, that is form the $A - \{a\}$ and $B - \{b\}$ sets and pick again new members. Repeat the whole process till one of the sets say A is exhausted and thus we indeed achieved an equivalence from A to a $B' \subseteq B$.
This sounds fishy today to someone who dealt with sets before beyond the level of Venn diagrams. But to even many mathematicians sounds okay.
I want to make a point now that is probably the most important in this whole article. To lament about the "ripening of ideas". I used to do this lamenting when the concept of sets in geometry or the decimal system as tool to see the irrationals is the subject. I always emphasize three things. My surprise about why these remained obstacles for so long, how then they become so easy to grasp when they are served on silver platter and finally that these silver platters hide some layers that then will be buried even deeper. But now the first time I will make a thought experiment.
So I'll present an argument that never has been made but could have been by Cantor. He must have felt that the gradual picking of pairs from A , B is not kosher simply because of the use of time. Time must be banished from all mathematical reasoning! Exactly the set concept made it clear that when a point sequence approaches a P point then this has nothing to do with time! Instead, the simple fact is true that in any surrounding of P there are infinite many points of the sequence. A stronger condition is if the sequence only approaches P. Then for any n natural there is an n depending surrounding of P where all sequence members after n are inside.
And surprisingly time can be avoided quite easily here to.
First of all we need a trivial but reoccurring crucial concept, the "proper" subset of a set. It simply means a subset not being the full set, so having still members outside.
So the big idea is simply regarding two functions, a g in A and a h in B that predetermine the possible choices. More precisely, g must be defined on all possible proper S subsets of A while h on all possible proper T subsets of B.
And they could give the next choices. So if S is already obtained in A we chose next from A the $g(S)$ member outside S and if T is already obtained in B we chose next from B the $h(T)$ member outside T.
All we need now is truly just a single first a choice from A and b from B.
The machinery of the g , h functions will do the rest. A beautiful self determination will produce the widening subsets, we are not involved.
Unfortunately, we can have a counter argument.

Namely, what if both g , h predeterminations last forever. So in spite of achieving bigger and bigger subsets of the two sets, none of them gets exhausted.
If for example one of the sets is the natural numbers then we might easily arrive after infinite many pickings to a still infinite set. Say we picked all the odds then all the evens are still left. Then again we might pick only every second and so again are left with infinite many and so on. In fact, in spite of that we visualized these very improbable exact choices it is still most probable that an infinite picking sequence should leave infinite many still there!
Luckily, a rescue counter counter argument can be made too.
Namely, observing that the picking that were solely for the purpose of making pairs could be done just in one set for the purpose of ordering it.
So what if we can show that with forgetting the simultaneity completely, in any single A set such full ordering of A can be obtained by the g prechoice function. Then imagining a same in B we have two ordered sets that then obeying the order can be assigned till the shorter is over.
So now we encountered the new problem of having definite shorter longer versions.
And this new problem of what kind of shorter or longer orderings could dictate a comparing of sets was solved. So in a sense the time usage was eliminated.
They even found a new name for such orderings as well-ordering. In fact, König thought he found a proof that the set of all decimals can not be well-ordered.
So when the error was realized then even more clearly the goal became how to well-order an arbitrary set. And so the bigger idea behind, the self widening by a function got buried. When Zermelo presented his proof that all sets can be well-ordered he emphasized the new axiom that allows such well-ordering of any A set.
But amazingly, the new axiom is only responsible for the “any A set” part.
A less universal existence of well-orderings can be proven without the new axiom.
This article goes back to this forgotten and never explored part and thus making the new proof itself finally crystal clear. But before we embark on this road we must mention an other problem with the orderings in general that has never been raised.
An order is visual only through the left and right on a horizontal line. But it can be shown with a heuristic generalization of Cantor’s fundamental result about the non sequencability of the decimals that: For any set we can create a bigger one.
Namely, the set of all S subsets of A , that is the $P = \{ S \,;\, S \subseteq A \}$ power set of A is always bigger than A.
With our earlier mentioned simplified criteria this means that by no f function can an A set cover P. That is: For every f function there is some $S \subseteq A$ not in $f[A]$.
And indeed, let’s regard the $S = \{ a \,;\, a \in A \text{ and } a \notin f(a) \}$. This can not be in $f[A]$.
Suppose it were, that is for some $s \in A$ we had that $f(s) = \{a \,;\, a \in A \text{ and } a \notin f(a)\}$.
We show that both scenarios of $s \in S$ and $s \notin S$ are impossible because they imply each other.
$s \in \{a \,;\, a \in A \text{ and } a \notin f(a)\} \;\rightarrow\; s \in A \text{ and } s \notin f(s) \;\rightarrow\; s \notin f(s) = \{a \,;\, a \in A \text{ and } a \notin f(a)\}$.
$s \notin \{a \,;\, a \in A \text{ and } a \notin f(a)\} \;\rightarrow\; s \notin A \text{ or } s \in f(s) \;\rightarrow\; s \in f(s) = \{a \,;\, a \in A \text{ and } a \notin f(a)\}$.

There is an interesting counter counter argument to our indirect reasoning.
Namely, we can say: What if there is no $a \in A$ at all that $a \notin f(a)$?
Then our S is the so called empty set that we regard as subset in any A set.
We’ll talk about this later again. But observe that even if we would try to avoid this empty set as a phantom, the impossibility of f having all subsets of A in its range still follows with an even stronger plausibility.
Indeed, if for all a we have that $a \in f(a)$ then we can make a simple branching in our argument. Either there are $\{a\}$ singular subsets of A that are not assigned to some members and so we are finished, or all these are assigned and so must be assigned to a and then we exhausted all a members and so no other kind of subsets of A are assigned at all.

This still can leave us pretty puzzled and so we show a follow up argument.
But first let's repeat what we did.
By the weak comparability of sets, an impossibility of $B \leq A$, that is of $f[A] \supseteq B$ being true for some f implies $A < B$. And this is what we showed, for our case as B being the set of all subsets of A that is being the P power set of A.
So we showed that for all f functions $f[A] \supseteq P$ is impossible.
Which also can be said as $P - f[A]$ being never empty.
We used the "anti diagonal" S subset of A as an example in the $P - f[A]$ set.
This is what made us so reluctant to see this as proof for $A < P$ since we showed only a single member in $P - f[A]$.
To make us convinced about our result intuitively too, we can introduce $A \ll B$ with the intended meaning that A is much smaller than B.
Namely, by claiming that for any f function $B \cap f[A]$ is smaller than $B - f[A]$.
That is, for any g function $g[B \cap f[A]]$ doesn't cover $B - f[A]$.
Now comes the amazing fact that if B is infinite then: $A < B \rightarrow A \ll B$.
The negated version means that if for a B infinite set there are f, g functions that $g[B \cap f[A]]$ covers $B - f[A]$ then there is a h function that $h[A]$ covers B.
As a lemma, we claim that if in a B infinite set an S subset covers $B - S$ by some function then S can cover the full B too by a k function. Then in our case too, for a k function $k[B \cap f[A]]$ will cover B and so $h[A] = kf[A]$ will cover B.
To prove our lemma, we must split S into two "equal" halves.
One half can cover S the other $B - S$ and thus S the whole B.
This was a very long detour to show that there are sets bigger than all points of a line and so the visuality of the orderings is questionable. And yet it was never really questioned and we'll come back to their abstract treatment in Appendix 4.

Zermelo

The problem of comparing sets and their possible orderings on their own were not gradually solved rather solved all in a complex bundle.
This article is about cutting this bundle in two by realizing that by going just a little bit beyond the immediate goal of the full well-ordering by a suitable g picking function we can claim a new theorem that already solves the avoidance of time by creating a well-ordering instead. Of course, this sounds nonsense!
How can I order if I don't have the crucial g function?
Quite simply, first regarding any g which might not guarantee a full ordering of A.
So stepping away from the A set to be ordered. But then what will be full?
The fullness will be decided by g itself. Namely, as a collection by g where g is not defined anymore. But the crucial point is the word "anymore". So it's not just the total being a set where g is not defined that is important but also that "before" through the collection process g was always defined.
If you pour water into a labyrinth it will get flooded till the first section with a hole.
So it's not the section having a hole and thus being non fillable that matters.
The water finds the first hole by itself. Similarly, a full self growth of any g function is that we will prove. This then using the suitable g in an A set given by the new axiom will instantly show that the full A must become ordered.

But before we embark on this road I want to sketch why the new axiom was so hard to realize. In spite of mathematicians using it earlier unconsciously.
The very plausible spatial collection used as the magic wand was the $\{ x \,;\, P(x) \}$ property collection. A totally unrelated problem with this was the famous Russell paradox. Using the same simple argument we showed above for the power set it turns out that the $\{ x \,;\, x \notin x \}$ collection is contradictory!

The solution became to restrict the applicable $P(x)$ properties.
But Zermelo recognized a totally different problem, namely that sometimes we don't want to collect all possible P sets as our new set only chose a sample from different properties. We can not give infinite many properties at once but actually all sets are regardable as properties. Thus all sets are actually a set of properties.
The s members of an S set are all different properties and so we can go into the third deeper level and pick a member from all the s members and form a sample.
There are some glitches of course! What if some s has no members, what if they have common members so we would pick the same, and so on.
So a better version is to pick the members but don't let them melt into a sample set rather remember which was picked from which s and so create a $c(s)$ choice function as sample function.
An example can now demonstrate the hidden assumption needed for full exactness in an argument that would have been regarded as trivial before.
We have a set of points on a horizontal line so that there is no leftmost point in it.
The points don't have to go to infinity because they can also approach an L point from the right so that L is not in our set. The almost trivial claim is that there is a P_n sequence of points in our set that goes to the left. So it might go to infinity, it might approach L or not, these are totally irrelevant. Only that P_n goes to left.
We might say: Pick a point as our first sequence member. There still has to be point in our set because there was no leftmost point in the set. So pick one again. Repeat this and we got our sequence. Now let's see the correct argument:
Let $G(P)$ denote the function that orders for every P in our set, the points being left to P in our set. By our assumption, every $G(P)$ value is a non empty set.
Now let's pick one point from each $G(P)$ set and let the function assigning these to the P points be $g(P)$. So for every P in our set $g(P) \in G(P)$.
Let's pick a single P_0 point from our set and let's call a set of finite many points in our set as a (P_0, g) collection if P_0 is in it and for every P member, the one next to its left is $g(P)$. Easy to prove that these (P_0, g) collections are widening sets.
Their combined set is infinite but every member has finite many member to its right.
So this finite number as index makes the combined set a sequence going left.
Of course, to fully appreciate what we achieved you must know a bit of Logic!
The singularly picked P_0 was okay!
If something exists we can make a named individual for it.
But to do this for infinite many existences simultaneously is not granted by Logic.
This is the missing capacity replaced by sets and Zermelo's New Axiom Of choice.
So the new axiom is actually a Logic axiom and thus the final seal on how Logic and Set Theory is intermingled.
For the $G(x)$ that had always non empty sets as values, we assumed a $g(x)$, a sample function for $G(x)$. That was our form of the Axiom Of Choice.
It's more sophisticated than a simple sample function picked from a set's members.
And we'll use later in Appendix 5. an even more sophisticated version where we allow that $G(x)$ has empty values where we simply don't pick members.

Our argument was very misleading in the sense that the Axiom Of Choice was our start. And this mistake is with us since the Well Ordering Theorem was first proved!
The amazing truth is that the Axiom Of Choice should only be used as a final step!
We will use an arbitrary f function to derive the essence of the proof that will be called as the f-widening Theorem. And then truly just as a punchline, we will use the Axiom Of Choice to find an f with which our f-widening theorem implies the Well Ordering Theorem. But if this is not enough, it turns out that this f-widening concept with arbitrary f functions has a wonderful visual meaning on its own.

To connect our very plausible later ideas to the comparability argument better, we must go into a further change that happened after Zermelo initiated the axiomatic description of Set Theory.
It became clear that this theory is less of the collections and more of the build-ups.
What we start to build from is almost irrelevant. And this created the strange consequence that this "theory of everything" became the "theory of nothing".
The "of nothing" was of course a bit of a stretch and a more appropriate wording could be the " build-ups from nothing". Namely, from the empty set.
This is the set that has no members. But why is it "the"? Because we want a theory of build-ups and so all the seemingly atomic or urelemental objects like numbers or points will be replaced by sets built up from the single 0 empty set.
And exactly that's why I will not use alternate fancy symbols for it rather the numeral form that later is also this same empty set.
There is also a very simple practicality in this singular empty set concept.
Namely, that we can define the equality of sets by simply saying that they have the same members. $S = T$ if and only if for every e set: $e \in S$ if and only if $e \in T$.
But we allowed something very strange also! Namely, that our single 0 will be a subset of every set! Indeed, by this equality definition it is obvious that being a subset should be defined claiming only one directional implication, so:
$S \subseteq T$ means that $e \in S \rightarrow e \in T$ is true for all e.
Now, if there are no e sets that are members of S because S is 0 then this implication is actually true for all e and T. So $0 \subseteq T$ is true for all T.
How can one thing be inside everything? Well, the empty set is not a "thing". It is the empty "build-up". And so everything else should also be treated as a build-up.

Two Moments In Time

Now this is a very strange sub title! Especially after emphasizing that time must be banished from math. And indeed, now I want to reveal something almost mystical.
Time can not be banished from anything! It is actually the essence of everything!
The title refers to now that I am typing this article in 2022 and to the moment when Cantor made his thought experiment how to compare sets.
The bridge between, one and a half century is certain. Because it became a past.
Poor Cantor went through hell when God has handed him the greatest abstraction in our short history. And still an unexplainably late present.
How could Euclid miss the concept of sets? He laid down his postulates about space as well as it was possible without sets. Hilbert the hero who stood out for Cantor in his 1900 math congress speech said that no one can take away the paradise that Cantor created for mathematics. And he not only defended Cantor but went back to Euclid and re-axiomatized Geometry. Grundlagen Der Geometrie is a monstrosity on one side that never be dealt with in high schools where the banishing of Geometry is happening anyway but on the other side it is a proof of that the postulates of Euclid were correctable to a perfect Logic. But now I have to be the pessimist.
Why could not see Hilbert that something much more recent hid a much deeper stuff to be perfectionized. This was the banishing of time by Zermelo from the crucial thought experiment of Cantor to list the two sets element by element till one gets exhausted and thus becomes equivalent to a subset of the other. And at this exact time Einstein made his thought experiments to realize that time must enter physics.
This article shows that it's not the spatial Axiom Of Choice that makes Cantor's exhausting assumption plausible. An unnamed theorem, the f-widening Theorem does that without any new axiom. And this clarification is part of a bigger picture about time that will finally shed some light on the full comparability of sets.

Widenings

So finally we turn to our plausible road to exactify Cantor's idea to compare sets.
The element by element ordering is regarded as element by element widening.
In fact, generalized as any widenings. This is logical because the subset relation is already given and very visual. But there is a much deeper logic behind this approach too. Namely, as a generalization of the induction among the naturals.
The simplest is to derive something about all naturals that inherits from number to number. A more versatile version, called complete induction uses inheritance from all the numbers under n to n. And an even more versatile version could be to make even the claims about such beginning sections of the naturals and to inherit those step by step to larger beginning sets. And this dictates the really big jump to continue the idea and make inheritance to wider and wider arbitrary sets.
So we will widen an S_1 starting set step by step through S sets that together as a W set will be called our widening. The word "chain" jumps in mind and it is indeed an accepted name for any W set of sets if for any two S , T members one is subset of the other. But this chainness is not enough to call such set a widening from S_1 even if we claim that S_1 is the narrowest member of W.
Simply because sets can infinitely narrow not just at the left end of a chain which assuming S_1 as minimal can be excluded easily.
Indeed, imagine a narrowing sequence of sets and placing an even narrowest S_1 inside the whole sequence. This remains a chain with the minimal S_1 but from this S_1 we can not gradually get to the other members.
Okay, so let's exclude any possible narrowing sequences!
That's right! This would indeed be the simplest definition of a widening but making much harder to prove our fundamental theorem. So we want a more positive definition that resembles Cantor original idea of the gradual widening.
And such is possible by saying that every "beginning section" in our widening has a next wider member. The problem is that this "beginning section" is paradoxical.
Indeed, let S be a member or as we will also call them a "stage" in our yet to be defined W widening. The T stages narrower than S , that is $W(S) = \{ T ; T \subset S \}$ should be the beginning before S. And this is perfectly correct but fatally useless.
Simply because our very axiom of a widening should be that every beginning has a next wider member. And that is exactly S for W(S). So using S already in the definition makes the restriction obsolete. We need a trick. And it was figured out by Dedekind and that's why he is the second person mentioned in our title.
This whole didactical situation is never emphasized, exactly due to the stupid use of the final well-ordering concept. Which is actually an extreme version of our above negative idea to exclude narrowing sequences. Namely, to exclude any Q subsets that have no narrowest member but saying it in the positive sounding form that:
Every Q subset has a narrowest member! Very simple and tempting but no thanks!
We want the hard yard! So here it is, the most fundamental definition:

D A V subset of a W set is subset-complete or in short complete in W if:
For every $S \in V$, no $T \subseteq S$ can be member in W outside V , that is in $W - V$.

I used here already the letter W though this concept of the completeness is totally general, meaningful and actually very very heuristic for an arbitrary W set.
It is the perfect preparation for our next widening definition.
In fact, we can already feel a timely aspect in this completeness concept.
V is the "past" collection while the outside $W - V$ part of W is the "now and the future". So the completeness means that no subsets of sets "already" collected can be collected "later".

Observe that the 0 empty subset is always complete since it has no S member.
This is very good! It means that by our next rule it can be always the start without being actually a member so a stage.
But observe the most trivial fact that W itself is a complete subset of itself too.
And this seems to contradict our very aim to define the beginnings as these.
But actually this is very good too! It just brought out the essence!
Namely, that only the complete proper subsets of W should be the beginnings.
So now we can be quite exact.

D

A W set is a widening if:
For every V complete proper subset of W there is a $V' \in W$ that:
$S \in V \rightarrow S \subset V'$ and $T \in W - V \rightarrow V' \subseteq T$.

So V' is a separator of the two sets V and $W - V - \{V'\}$. It is the present.
Wider than every stage of the past and narrower than every stage of the future.
The actual S_1 starting stage of a W widening is 0' unless W is itself 0. Indeed:
Observe that the 0 empty set abstraction makes 0 itself a widening already.
Because it has no members and so our requirement is true for "all of them".
We can make the V' separator member more abstract by introducing two generalizations. From the combining of two sets as: $S \cup T = \{e \,;\, e \in S \text{ or } e \in T\}$ to the content of a V set as $\bigcup V = \{e \,;\, e \in S \in V\}$. So actually combining all S members of V. From the common part of two sets as $S \cap T = \{e \,;\, e \in S \text{ and } e \in T\}$ to the full intersection of a V as $\bigcap V = \{e \,;\, S \in V \rightarrow e \in S\}$. So actually being the common part of all S members of V.
With these, the claim about the V' separator is that: $\bigcup V \subseteq V' = \bigcap (W - V)$.
But this form is a bit confusing because the left side makes it seem as if V' could be a member in V which of course is impossible by the right side.
In truth the left only shows a double duality! Namely as first that $\bigcup V$ can be a member in V and then definitely $\bigcup V \subset V'$ or $\bigcup V$ is not a member in V and then having the duality that V' can be $\bigcup V$ or something even wider. We'll always chose this second so use wider V' next stages after a limit V than $\bigcup V$.
Which is logical because for non limit V this must be done for sure.
But the more important versions of expressing our requirement about V' are with the earlier introduced W(S) beginning concept.
Firstly, as all V complete subsets being beginnings because: $V = W(V')$ and secondly as all stages being separators because $S = W(S)'$. Which also implies that for every two S , T stages $S \subset T$ or $T \subset S$ also called as W being a chain.
Now we turn to the most important idea, namely using an f function to predetermine the possible widenings. But what should f be defined on?
The simplest would be the V complete proper subsets but as you would expect it by now, it will be the content of them $\bigcup V$.
Also observe that these are only stages in W if $\bigcup V \in V$ already, so by our mentioned convention only when V is not a limit beginning.
An f is a widening function if for every S set where f is defined: $f(S) \supset S$.
Such f will widen by itself from any S_1 starting stage until it reaches a stage where f is not defined anymore. Strangely, this heuristic self growth idea will not be apparent in our next definition. The reason is simple. Because it will not be in it!
It will not define this grand idea rather the much more humble "partial" f-widenings and these are that we simply call as f-widenings. The grand idea will only come back in how we use these partial ones in our f-widening Theorem. Namely, we'll combine these "partial" f-widenings. The reason for this approach is simply that this is the easiest. Nevertheless in Appendix 2. the direct approach to define the final or total f-widening will be explained showing its heuristic beauty. But now instead:

D

Let f be a widening function
A W widening is an f-widening if every V' separator is $f(\bigcup V)$.

A least important advantage of using an f function is that the empty start becomes even nicer: $f(\bigcup 0) = f(0) = S_1$ is the real starting stage if f is defined on 0.
So if you want to use a "real" S_1 starting stage widened by an f "real" widening function then just extend f to be defined on 0 as S_1.
Finally, the most important definition that allows to express the mentioned success about the total f-widening is this:
An f-widening is continuing if f is defined on $\bigcup W$.
Indeed, the success is the opposite, a non continuing f-widening.

T

f-widening Theorem:
For any f widening function there is a unique non continuing f-widening.
There is a more informative version as two double theorems:
Let $\mathbf{W}$ denote the set of all f-widenings. Then:
a. $W, W' \in \mathbf{W}$ and $W \neq W'$ $\rightarrow$ $W' = W(S)$ or $W = W'(S')$.
b. $\mathbf{V} \subseteq \mathbf{W}$ $\rightarrow$ $\bigcup \mathbf{V} \in \mathbf{W}$.
Which have the immediate consequences:
A. $\bigcup \mathbf{W}$ is a non continuing f-widening.
B. This is the only non continuing f-widening.

P

It's useful to emphasize right here at the start how the simple but ingenious concept of the complete subsets make the proof work. It will allow a two directional inheritance. One from sets to their combined total and one backward from a total to the members. The first will be used for proving a. while the second for b.
But to be even clearer and showing the finesse already in how we stated our theorem as two leveled, we first show the immediate consequences:
To see A. first observe that by b. $\bigcup \mathbf{W}$ is an f-widening.
Then observe that if $\bigcup \mathbf{W}$ were continuing, that is f were defined on $\bigcup\bigcup \mathbf{W}$ then $\bigcup \mathbf{W} + f(\bigcup\bigcup \mathbf{W}) = \bigcup \mathbf{W} \cup \{f(\bigcup\bigcup \mathbf{W})\}$ were a wider than $\bigcup \mathbf{W}$ widening so $\bigcup \mathbf{W}$ were a proper subset of this, contradicting that all f-widenings are members of $\mathbf{W}$ and thus subsets of $\bigcup \mathbf{W}$.
For B. observe that by a. any other than $\bigcup \mathbf{W}$ f-widening is a beginning in $\bigcup \mathbf{W}$ and so is continuing.
The proof of a. has a heuristic twist.
Above we said that the inheritance of the completeness to combinings will be used:
In any W set a combining of complete subsets is also complete in W.
This is indeed trivial by the definition of completeness regardless of widenings.
The twist will rely on an other trivial inheritance, true only in widenings but still not depending on f-widenings. Namely, that for a V complete proper subset the V + V' set is a wider complete subset. And now the twist:
Let's regard the two different W , W' f-widenings of our claim and let's combine their common complete subsets into the V set.
By the inheritance to combinings, this V is either the full W or a W(S) beginning in W and similarly either the full W' or a W'(S') beginning in W'.
V = W = W' can not be the case since we assumed two different widenings.
But being beginning in both, that is V = W(S) = W'(S') is also impossible because then by W , W' being f-widenings, $S = S' = f(\bigcup V)$ were and so V + S = V + S' were a wider common complete subset than V by the inheritance in widenings.
So the only option left is that V is a beginning in one and the full in the other.
Which is exactly what a. claims.

For proving b. we must show that the fact of the V_W complete subsets being separated by $f(\bigcup V_W)$ inherits from the W members of **V** to the $\bigcup\mathbf{V}$ combining. So let V be a complete proper subset of $\bigcup\mathbf{V}$. This implies that there has to be some S^+ member of $\bigcup\mathbf{V}$ that is wider than all members of V.
We claim that V being a complete proper subset in $\bigcup\mathbf{V}$ inherits backward to any W^+ in which S^+ is a member. The subsetness inheriting means $S \in V \to S \in W^+$.
Enough to show that for every W that S came from: $S \in V \to S \notin W - W^+$.
If $W - W^+$ is not empty at all then by a. W^+ must be a complete subset of W.
So since $S^+ \in W^+$ and $S \subset S^+$ thus S can not be in $W - W^+$ indeed.
The properness inheriting is trivial by S^+ and so all boils down to the mentioned heuristic backward inheritance of the completeness:
A V complete in $\bigcup\mathbf{V}$ and being inside a $W \in \mathbf{V}$ is complete in W too. And indeed: If an $S \in V$ had a $T \subseteq S$ being in $W - V$ then S would be in $\bigcup\mathbf{V} - V$ too**.**
So V is a complete proper subset in W^+.
There the separator stage of V is $f(\bigcup V)$ and so this will come into $\bigcup\mathbf{V}$.
Now we only have to show that it will remain a separator of V in $\bigcup\mathbf{V}$ too, so:
$S \in V \to S \subset f(\bigcup V)$ and $T \in \bigcup\mathbf{V} - V \to f(\bigcup V) \subseteq T$.
Both are trivial if S , T are in W^+. And S has to be such for sure. So only the second must be shown for T outside W^+. Suppose it came from a $W - W^+$.
Using again a. W^+ is a beginning of W, so all members of W^+ are subsets of all members of $W - W^+$ in particular $f(\bigcup V) \subseteq T$.

Elemental widenings, The Axiom Of Choice.

Now we specify our f-widenings to be elemental, meaning that S_1 is an elemental {s} starting stage and then also every V' is elementally increased.
The $f(\bigcup V)$ widenings will now be $\bigcup V + g(\bigcup V) = \bigcup V \cup \{g(\bigcup V)\}$.
So this g function that orders not new stages rather new elements to the stages is giving the widenings. $S_1 = f(0) = 0 + g(0) = 0 \cup \{g(0)\} = 0 \cup \{s\} = \{s\}$.
So simply having $s = g(0)$ gives the start.
All this hocus pocus becomes nicer with finally turning to Zermelo's punch line.
So now we are among the subsets of an A set and we claim that a c choice function can assign or "chose" a member from any S subset: $c(S) \in S$.
This is the exact opposite what our g elemental widening function should do.
This kind of prefect oppositeness must mean a match made in heaven!
And indeed, let's regard g(S) as c(A – S).
So the full picture is: $f(S) = S + g(S) = S + c(A - S)$.
This will always define a next, one element wider subset of A.
The above just tricky defining of g on 0 now becomes very natural too.
Indeed, $g(0) = c(A - 0) = c(A) = a$ can be our starting element.
Also observe that our g is defined for all proper S subsets of A but not for A.
Indeed, then $A - A = 0$ and so c can not pick a member.
It's mind boggling how a complicated f continuation function becomes a single widening by simply starting from the chosen a starting member.

But all this doesn't reveal why this elemental widening in an A set is so important.
It's twofold! Firstly, that it is not merely in A but actually on A and secondly that the same can be done on an other B set too.
The "on A" , that is a W with $\bigcup W = A$ follows from our f-widening Theorem and the used Axiom Of Choice.

Indeed, if W is the claimed non continuing f-widening with the used g and c then $\bigcup W$ can not be a proper subset of A because then W were still continuing.

The usual vision is that such elemental full widening on A actually well-orders A. Quite simply as regarding $a' \prec a$ when a' is in the beginning before a.
This well-ness however isn't the beautiful meaning that all Dedekind beginnings, that is complete subsets have a next member, rather something else that we come to later. The beautiful beginning meaning would make sense as replacement because doing the same on a B set too we indeed visualize them as practically ordered the same, except one being longer and the other shorter. And indeed this is the usual way to continue toward the whole start of our journey, the comparability problem.
Namely, they introduce the concept of similarity which is simply an equivalence that follows the orders. And even more importantly can be defined by the order.
Then it can be shown that one of them is similar to a beginning of the other which at once implies an equivalence with that beginning too.
Unfortunately, we must redo a lot of steps of the f-widening Theorem! So this approach is ugly in some respect and yet crucially important from an angle that is never even mentioned. Namely, that the particular f-widening becomes irrelevant.
Having elemental widenings on A and B implies the comparability.

Appendix 1.: A grander vision to solve comparability

Obviously if A and B have common members then those could be ignored because then an equivalence from one into the other without those would at once mean one with those too. So now A and B are disjoint. Also, two (a , b) , (a', b') ordered pairs are disjoint if they have no common member.
So an E equivalence in-between is simply a set of ordered pairs with first members from A second from B and every two pairs being disjoint.
An E is maximal if all A members appear as first member in the E members or all B members appear as second. The existence of such is that we must show.
We can use a totally opposite approach as above. Combine the f-widening Theorem with the Axiom Of Choice and see why a maximal equivalence must exist.

For any non maximal E equivalence we can pick new a , b members from A and B and define an f(E) widening function as the incremented equivalence with adding (a , b) to E. Then for the claimed non continuing W f-widening $\bigcup W$ will be a maximal E. Very visual but observe that beside the incrementation giving a new equivalence we also used that the $\bigcup W$ combinings remain equivalences.
Also observe that with this suggested vision I did the exact opposite that I was championing for, to separate the f-widening Theorem from the Axiom Of Choice.
Still, we achieved something general as:

D A U set is widening closed if for any W widening subset $\bigcup W \in U$.

T Elegant Theorem:
If U is widening closed and an f widening function's domain and range are both in U then the domain can not be the full U.

P Let's regard the **W** set of W f-widenings. All W-s and $\bigcup \mathbf{W}$ too are subsets of U.
$\bigcup\bigcup \mathbf{W}$ is member of U but it can't be in f's domain since $\bigcup \mathbf{W}$ is non continuing.

This raises the problem of where this result lies. The most general result about finding a set where an f widening function is not defined is this:

The range of f combined can not be in the domain.
Indeed, otherwise the f value on it were wider contradicting that we combined all.
So this is the triviality that is refined by the f-widening Theorem and the above too.

Appendix 2.: The f-closed approach

Combining partial f-widenings was necessary because the existence of a directly described total is usually even harder. Yet now we show such alternative approach.
The concrete idea is to start with something more general than the W f-widenings.
Namely, Z collections of stages that could be called f-closed sets, meaning:
$V \subseteq Z$, f is defined on $\bigcup V \quad \rightarrow \quad f(\bigcup V) \in Z$.
There are three very surprising things here. The most evident but least important is that there is no mention of complete subsets rather any V subsets. The more subtle surprise is that the assumption of the f widening function being defined on $\bigcup V$ is not a requirement rather a condition and only the addition of the then existing $f(\bigcup V)$ is a requirement. And the least visible but actually the most important feature is that we do not have the proper restriction on $\bigcup V$ so it can be the full $\bigcup Z$.
This shows that we are not defining some partial widenings rather an instant relentless widening by f. We can even prove easily that any such f-closed set is non continuing, that is f can not be defined on a $\bigcup Z$. Indeed, suppose the opposite.
Our rule used for $V = Z$ then gives $f(\bigcup Z) \in Z$ and so $f(\bigcup Z) \subseteq \bigcup Z$ follows contradicting $f(\bigcup Z) \supset \bigcup Z$. But now comes the blast that an f-closed set can be easily widened in a negative sense too. So we can add obvious junks to it.
For example sets where f is not defined at all.
Luckily we have an again trivial feature that helps. Namely, that the common part of f-closed sets remains f-closed. And so it seems plausible that if $\mathbf{Z}$ denotes the set of all Z f-closed sets then the $\bigcap \mathbf{Z}$ common part of all of them should be the junkless non continuing full f-widening. The exact proof is very heuristic but requires finally:

T Q minimality or Well-ordering Principle: In a W widening:
$Q \subseteq W$, $Q \neq 0 \quad \rightarrow \quad$ Q has a narrowest member.

P Let V be the set of those members of W that are narrower than all members of Q.
This V set is actually a complete proper subset!
Indeed, it is complete trivially and it is not the full W if Q was non empty.
Thus by W being a widening $V = W(S)$. Also, due to the chainness of W:
A non Q member is either narrower than all Q members or wider than some.
The first ones are V but they are also all the narrower than S stages.
Thus S can not be wider than some Q member either. So $S \in Q$.
Finally, all other Q members are wider since they can not be among W(S).
Observe that even if Q is the full W and so $V = 0$, the rule of widening guaranties a narrowest stage as 0' unless W itself is 0 as we explained earlier.

T $\bigcap \mathbf{Z} = \bigcup \mathbf{W}$.

P Enough to show that $\bigcup \mathbf{W}$ is f-closed set and has no stages that are not in any other f-closed set. The first is quite easy by introducing the concept of completion.
In a W widening a V subset's completion denoted as W[V] is the set of all those stages that are subsets in V members: $W[V] = \{ S ; S \in W \text{ and } S \subseteq T \in V \}$.
Now observe three things:
W[V] is complete in W , $\bigcup V = \bigcup W[V]$, $W[V] \subset W \rightarrow \bigcup V \neq \bigcup W$.

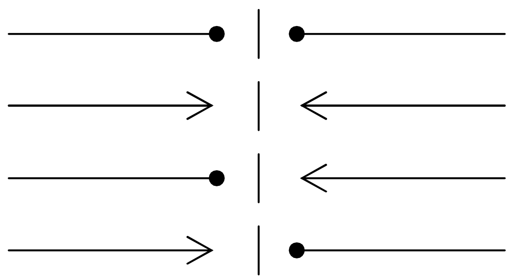
Thus in a W f-widening since f is defined on all complete proper subset's content, it is also defined on all $\bigcup V \neq \bigcup W$ and since $f(\bigcup V)$ is member in W the closedness is satisfied for all V for which $\bigcup V \neq \bigcup W$.
In the case of $W = \bigcup \mathbf{W}$ we know that f is not defined on $\bigcup W = \bigcup \bigcup \mathbf{W}$ and so here the closedness is true for all V and so this is an f-closed set.

Now we show that all stages in $\bigcup \mathbf{W}$ are in every f-closed set.
This actually means that all stages in any W f-widening are in every f-closed set.
Suppose there were some stages in a W that were not in every f-closed set.
Let the set of all these be the Q subset of W. By our assumption Q is not empty.
Thus by the well-ordering principle Q would have a narrowest S member.
To see that the assumption of a non empty Q was a contradiction enough to show that S must be in all f-closed sets.
And indeed, the W(S) beginning is subset in every f-closed set since S was the first stage not being in all of them and thus f(W(S)) = S is in all of them too.

R Now we venture into two exciting continuations of the Q minimality.
The first is an even more abstract counterpart, a P maximality.
This is the principle of transfinite induction.
So P which is again a subset of W can be regarded as certain property that we can prove to be true for all members of W if we can show it to be true for the first stage and show the inheritance from the V complete proper subsets to the next stages.
The beautiful abstract argument relies merely on the two assumptions of the Q minimality and that every stage is the first one or the next one to a V.
Indeed, then suppose that P is inductive and yet some stage were not P.
There would have to be a first such S and then the V beginning before had only P members and so by the induction inheritance S would have to be P.
Exact copy of what we do among the naturals.
Now comes a second, quite opposite, not abstract but deep dialectical stuff.
Actually we mentioned it already! As the simplest possible rule for the widenings.
It says that there are no narrowing sequences in W. Deep stuff. Indeed, we see that all the huge infinites can be well-ordered and yet we can go backward in any of these only finite many steps. Anybody who says "what's the problem" is an idiot.
But it gets much deeper! Because though formally we can "see" that the Q minimality implies this, even this step is non plausible. Indeed, the Q minimality implies that a sequence as particular Q must have a narrowest member too but this still doesn't make the impossibility of the infinite narrowing evident.
Only our logic says that if in a sequence of sets there is a narrowest then going through them, in finite steps we arrive at this narrowest member and then all the others must be inside this and so we refuted the infinite narrowing.
Again, if you don't see my point and say "what's the problem" I give up.
We are different species. (And that is not a joke actually!)
But now comes the real point that even "you" can not deny to be very strange.
Namely, that this particular case of Q being merely a sequence implies the full Q minimality but! And now comes the weird stuff, this requires the Axiom Of Choice.
Indeed, the argument is actually the same we used in our initial problem on a horizontal line but now to show the indirect version of our present claim that in any backward infinite set there is a backward sub sequence.

Appendix 3.: Dedekind Cut, the first recognition of complete subsets

This connection is never mentioned in the literature and there are historical reasons for this amnesia. Namely, that when Dedekind realized the inner definability of the beginnings, he used it for something else than we did. Namely, to define continuity.
Cantor had his own definition, as the so called “common point principle”.
Seemingly a very logical approach, locating the common point of shrinking intervals. Continuity, that is not missing any point on the line means that we claim the existence of such common point. The problem as always, lies in the details.
First of all, narrowing intervals do not always shrink to a common point!
Only if they are closed intervals. Indeed, even half closed intervals can narrow down to the missing end point and so the common part of the intervals will empty out!
This is so trivial that we may think the assumption of closedness is obvious too.
But this is not so because open intervals will narrow to a common point if we do not allow them to narrow to the end points. So after all, the closedness is not the essence! And then if you know the continuation as the Common Point Theorem, then it turns out that the essence is still the closedness.
By the way, you can find that theorem in the “Topology 101” section of my “Measure In The Unit Interval” article.
But Dedekind found an alternative way to define continuity through the ordering of the line. In fact, he realized something more general for any ordered set where a left and right is defined for the members.
The big trick is hidden in the quite visual naming of his idea as the “Dedekind Cut”.
So we don’t focus on the members rather the subsets. Namely complementing ones.
In other words, we separate our set into two complementing subsets.
He added to this separation something ridiculously simple yet powerful.
Namely, that every member of one side is left from every member of the other side.
So the word “cut” is very appropriate if we realize that the cutting, which will be our little vertical line is not a member of the set and the two halves L and R are even pulled apart:



Quite amazingly, these four scenarios are all that is possible for a pair of cuts.
The first is the “gap” where L has a right most member and R has a left most.
The second is a “hole” where neither side has end points only approachings.
The third has right most left side but approach from the right, the fourth is in reverse.
We obviously feel that the first two should not be true about a continuous line!
So the definition of continuity is simple!
It means having only third or fourth kind of “cuts”, that is possible separations.
The natural numbers have only gaps, the fractions can have any cuts except gaps.
And our subject, the well ordered sets have only gaps or the last left approaches.
Combining these two is easy by saying that the right cut has always a first member.
And thus we indeed arrived at the well-ordering if we add that the full set itself must have a leftmost member too. Or as I prefer it:
We arrived to a spatial abstraction of the “transfinite time” concept in which:
The whole set has a start and every left cut as “beginning” has a next member.

Appendix 4.: Ordering defined

In these arguments above the $\prec$ ordering itself was taken naively.
To make it precise here are its three rules:

1. Asymmetry : $x \prec y$ and $y \prec x$ are never both true for any x , y.

The $x = y$ equality is part of Logic and thus already this simplest 1. claim implies two things: First that $x \prec x$ is never true.
Indeed, in 1. we must allow replacing y with x so by claiming that $x \prec x$ and $x \prec x$ are never both true, we actually claim that $x \prec x$ is never true either.
Secondly, that if $x = y$ then neither of $x \prec y$ or $y \prec x$ are true for any x , y.
Indeed, we can replace equal members too and so either of $x \prec y$ or $y \prec x$ gives the other. Or to be truly a formalist, just replacing one of them we get $x \prec x$ or $y \prec y$ and we just proved that these can never be true.

2. Totality: One of $x \prec y$ or $y \prec x$ is always true for any $x \neq y$.
   To use formal Logic: $x \neq y \rightarrow x \prec y$ or $y \prec x$.

3. Transitivity: $x \prec y$ and $y \prec z \rightarrow x \prec z$.

In 2. we said "one of" which more precisely meant "at least one of" but using 1. we can be sure that actually "exactly one of".
By the way, the "at least one of" is usually denoted as $\vee$, while the "exactly one of" is abbreviated as ∇. Actually, 1. and 2. combined can be expressed as:
$x = y \ \nabla \ x \prec y \ \nabla \ y \prec x$ which is also called as the trichotomy.
We of course avoided ordered sets altogether because we used the widening sets and so we might think that thus all these ordering nuances became immaterial too.
But not so! Indeed, only asymmetry and transitivity follows from using the subset relation. Totality does not!
So for a set of sets it is not obvious at all that among the regarded sets:
$S \neq T \rightarrow S \subset T$ or $T \subset S$.
And indeed, this was our requirement of being a chain.

Appendix 5.: Relational maximality theorems

Zermelo presented his Well Ordering Theorem in 1904 and ten years later Hausdorff proved a new principle that for its proof required to redo all the steps.
But decades later the principle's trivial consequence was again rediscovered actually twice and again with long new proofs. All these were signs of the missing recognition that the f-widenings must be handled without the Axiom Of Choice.
The Axiom Of Choice is only the punch line of Zermelo's proof.
But this punch line was a bit narrow too. A natural generalization is to let our g widening function pick not from the $A - S$ subsets, rather from some $G(S)$ subsets of $A - S$. A further second generalization is to allow $G(S)$ subsets that are empty.
We can even avoid getting our g from a c by stating the Axiom Of Choice for functions: For any $G(x)$ function, there is an $g(x)$ function defined on all those x sets where $G(x)$ is not empty so that $g(x) \in G(x)$.
So $g(x)$ is a sample function from $G(x)$ where such sample member is possible.
Our particular $G(S)$ will be defined for all S subsets of A and having values as subsets in $A - S$. Including the mentioned crucial empty possibility.
Thus of course a termination will not automatically give A as content.
Instead, it will be an M subset of A where $G(M)$ is empty.

The most important third new idea is that this G(S) is defined by an R relation in A.
Namely, as those t elements of $A - S$ for which s ® t stands for all $s \in S$.
Observe that for an arbitrary R relation it is quite expectable that many s , t pairs are not related at all. But widening with a g chosen from G(S) we always get a W g-widening that in $T = \bigcup W$ all pairs are related for sure.
This is usually called as T being "total".
If in particular $W = \bigcup \mathbf{W}$ then $G(\bigcup W) = G(M)$ must be empty, that is:
There are no t elements in A outside M that s ® t would be true for all $s \in M$.
This can be called as M being "maximal" in A for the R relation.
And this is the whole point.
Without such maximality, a total T is trivial as any single s ® t pair.
So our most general result following from the f-widening Theorem is this

Basic Theorem:
For any R relation in an A set there is an M set in A that:
a. M is total, meaning that any pair in it, is relating in at least one direction.
b. M is maximal, meaning that there is no t outside M that s ® t for all $s \in M$.

By the way, the $\prec$ ordering defined in M by W is such that $s \prec t \rightarrow$ s ® t.
But this well-ordering came out as just an icing on the cake not emphasized.
All the next theorems follow from this not even named Basic Theorem step by step.
Which would make us think that they become more and more trivial consequences.
But this is far from the truth and that was the reason that they had to be discovered so slowly. So here with using the Axiom Of Choice in the background, somehow even the trivial consequences can become actually newer and newer smoke and mirrors.
The first step is to regard R relations that are partial orders having the following two rules:
1. a ® b and b ® a are never both true. Asymmetry.
2. a ® b and b ® c $\rightarrow$ a ® c. Transitivity.
Then the $\prec$ well ordering in M will be the same as our s ® t relation.
And so our above Basic Theorem becomes the Hausdorff Maximality Theorem:

For any R partial order in A there is an M maximal among those subsets of A where R is total.
Observe that if this M has a biggest m member then it is not only maximal in M but in the whole A.
Otherwise, that is having an m' bigger member in A, M weren't maximal.
And amazingly there is a very simple assumption that guarantees this.
Namely, that for any total S subset in which there is no maximal member, we must have a b member in A that is bigger than all members of S.
Indeed, then for the M guaranteed by Hausdorff, only the existing maximal member can be true otherwise we had a b outside and so M were not maximal.
The maximal member of an S or such outside b together are usually called as "upper bounds of S" and so we have the following trivial consequence of Hausdorff:

Kuratowski Zorn Lemma:
If in an A partially ordered set every S totally (well) ordered subset has an upper bound, then there is a maximal m member in A.

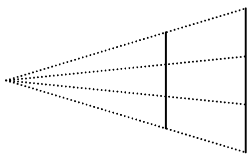
We may wonder if this artificially created condition of an upper bound existence could ever be useful. But I told that we are walking in smoke and mirrors.
In fact, the R relation itself can be regarded as the simple $\subset$ subset relation among the A members and still get a very strange condition that guarantees the upper bound condition of Kuratowski Zorn:

Teichmuller Tukey Lemma:
If in an A set $a \in A$ if and only if every finite subset of a is element of A, then there is an m maximal element in A regarding the $\subset$ relation.
In other words, an m that is not subset of any other element of A.

Let S be a subset of A that is totally ordered by $\subset$.
Enough to show that every finite $\{s_1, s_2, .., s_n\}$ subset of $\bigcup S$ is element of A.
Indeed, then by the finite membership condition $\bigcup S \in A$.
So there are upper bounds in A and thus by Kuratowski Zorn we get the claimed non continuable m element by $\subset$.
Now to see that indeed every such finite subset of $\bigcup S$ is in A, observe that these s elements came from some $a_1, a_2, ..., a_n$ elements of S.
Let the widest be a_i which then has all the $s_1, s_2, ..., s_n$ elements inside.
Then by the other direction of the finite membership condition: $\{s_1, s_2, .., s_n\} \in A$.

Appendix 6.: The Transfinite

I mentioned already that the left and right is questionable due to the existence of sets being bigger than the line where we see our left and right. In spite of this, instead of the widenings we accept the visually simpler $\prec$ relation.
We'll still use W, now standing for well-ordered set and V for subsets of these.
But first we must mention something mystical about our Euclidian ordering.
The projected pictures onto the huge screens of the cinemas are originating from a small sized film or nowadays from a digital media. And we all know that the pixels are visible on the TV screens too. So we might think that all this magic is due to the small size of atoms that allows to put all the millions of pixels onto a small surface.
But this misses something deeper that even physicists are taking seriously only lately.
Namely, the points! The Greeks were lamenting about them a lot.
Both just spatially and involving time. By the way, is time spatial too? Having singular points as the "nows"? I leave all these now.
Instead I want to show the single most important paradox of the Universe.
The paradox of the projection! A smaller interval is projected onto a bigger:



The connecting lines make a one-one assignment of all the points of the two intervals.
The magic is very trivial and unnoticeable. Namely, that the order is kept.
The left or right, or on our picture the up and down is kept.
One square millimeter has all the pixels needed for a screen of the whole Universe!
Somehow this miracle is still not connected to the miracles of the transfinite that we come now. Very probably making this connection will have to alter Set Theory.
The whole point of the widenings or as now we obeyed the inevitable, the well-ordered sets, is that for any two sets one must be similar to a beginning of the other. And so we could think that this is the end of the story. A monotone order allows everything to be part of it. Of course something should make us suspicious.
Namely, that the jump in set sizes should be connected to their well-orderings too!
The real beauty as always, is how logical and simple this connection is.
Indeed, what was the most important concept that lead us to everything beside the obvious membership relation? Being subset! And so the monotony of the well-orderings also hides a yet unused side of the subsets.

Even historically, the first deeper contemplation about the infinite orders was when poor Galileo using his unfortunate law of the fallings, sidetracked and noticed that the odds or evens are just as much as the full set of numbers.
His law was that the falling distances are the odd multiples of the falling distance in the first second. Had he realized that the odds always sum up as the squares:
$1+3 = 4$, $1+3+5 = 9$, $1+3+5+7 = 16$, . . then he could have regarded instead of the consecutive falls, the total from the top and say that these are the square multiples of the fall in the first second. But then we are not forced anymore to whole multiples and so quite simply the falling distance is proportional to the square of the falling time.
So we can calculate this distance for 3.75 seconds too!
But back to the odds and evens, these are not the simplest surprises either.
Cutting of an initial segment of the naturals will leave also a same infinite sequence.
By the way I believe that paradoxes have a strange underlying potentiality world by using different plausibilities intermingled. They are the smoke and mirrors.
A perfect example of this was the Achilles Paradox that intermingled the use of chasing someone with the much simpler finite sum paradox.
So I will now show a version of the mildest cut off paradox with an edge given to it.
By mildest I mean that we cut off the single first member.
So imagine beggars in a row each having a coin. I tell the first to give me his coin and take the one from his neighbor and tell him what I told him too.
Everybody gets back his coin and we made an extra coin out of nothing, or rather from the infinity of the coins. Soon we'll show a new angle of Galileo's observation.
But first something simpler that still concerns the beginnings. Indeed, we can make the simplest distinction of the well-orderings by the beginning cut off idea.
Simply, because sometimes such cutting off will not leave the same kind of sequence.
If we steal the number 1 from the A set of the naturals then the well-ordering of A in its natural increasing order remains similar that is "the same" for the $A - \{1\}$ set.
Let ω denote the well-ordering itself. If we place the stolen 1 after all the naturals then we get an $\omega + 1$ "type" of well-ordering. Of course, we could have stolen the number 2 just as well. The new order type would still be logically named as $\omega + 1$.
We can steal even more members and thus obtain $\omega + 1 + \;.\;.\; + 1 = \omega + n$ types.
Using Galileo's idea we can even turn A into an $\omega + \omega = 2\omega$ type well-ordering.
As we see, a whole arithmetic of these well-ordering types or so called ordinals is offering itself but we don't want to bog down in these trivialities.
The more important point is now that the original A set version is still there too.
In fact, we could have used new artificial members and leave A intact.
A very deep distinction that we still don't want to tackle now.
Instead, we just observe that whatever set representation an $\omega + n$ or $\omega + \omega$ order type, that is ordinal has, there it won't be true anymore that a beginning can be cut off with leaving the end the same kind. So these ordinals are not "cuttable" any more.
We might think that we grasped something deep but NOT!
Indeed, after ω the next cuttable ordinal is $\omega + \omega + \;.\;. = \omega\omega = \omega^2$.
So a simple sequencing as step generates these cuttable ordinals.
But we still feel that these are special due to being jammed towards the end.
The natural idea is to use something else that we already know as a big jump.
Well, Cantor recognized that there are non sequencable infinites and then generalized it as bigger infinites for of any given set. By the Well-ordering Theorem these can be well-ordered too and so there has to be a shortest well-ordering for every set size.
In particular, a first well-ordering that is not sequencable.
Since ω denotes the well-ordering of the naturals thus we'll denote this first non sequencable well-ordering type, that is ordinal as ω_1.
Then the same argument tells that there has to a be a first ordinal bigger in size than ω_1 which we'll denote as ω_2 and so on.

These ordinals that are bigger in set size too than all the smaller ordinals, are called as cardinals.
The combining of our previous sequence, also writable as $\omega_1 + \omega_2 + \ . \ . \ . = \omega_\omega$ will be again a new cardinal because it can not be equivalent to any of the members.
But we at once see a big difference between the members and the sum.
Namely, while this total has a simple ω like sequence as subset that goes all the way, at the members this is not apparent at all. And indeed, our hunch is correct!
Let's call a subset that goes all the way, so not contained in any beginning of a well-ordered set as a cofinalling subset. And let's call an ordinal that has no shorter cofinalling subset a cofinal. We claim that ω, ω_1, ω_2, . . are all cofinals.
While this is trivial for ω it is far from trivial for the rest of them.
In fact, it relies on a generalization of Cantor's other original recognition that the fractions are sequencable. Namely, on the fact that the A^2 set of pairs made from an A set's members is equivalent to A. This itself can be obtained by using the earlier mentioned boring monotony of the cuttable well-orderings.
Observe that we are already walking in the strange world of the transfinite!
We use sets in general like this A was above, also W well-ordered sets and well-ordering types, that is ordinals too, usually denoted by Greek letters, α , β , . .
We didn't define the precise shifts from sets to ordinals but they are always plausible.
For example, if $A^2 \sim A$ is proved then $\alpha^2 \sim \alpha$ follows at once.
Also observe that these transitions between sets and ordinals is inherent by the truths themselves as the above example shows. Indeed, as I said, the proof of the set form $A^2 \sim A$ uses the $\alpha^2 \sim \alpha$ form with regarding cuttability but next we'll use the set form to get the ordinal result that ω_1, ω_2, . . are cofinals.
In fact, we show that if α is a cardinal and the next cardinal is β then such "next" β cardinal is always a cofinal. Indeed, if β were not a cofinal then it had a V subset with a γ shorter than β well-ordering type. But since β is a cardinal, this also means that γ is not just shorter than β but smaller in set size too and so V would be a maximum α sized set of ordinals that are each maximum α sized too.
And this is impossible because then the combining of the V members could only be a maximum $\alpha^2 \sim \alpha$ sized set, contradicting that it is actually β .
There are two obvious questions after this grand result.
The simpler one is whether all cofinals are cardinals. And the more subtle is whether only those cardinals are not cofinals that are not next ones, that is are limit cardinals.
Before we prove the first we mention two trivial consequences of being a cofinal.
The first is that it implies being cuttable. Indeed, if a cofinal were not cuttable then cutting off some beginning would leave a smaller end but this were a smaller cofinalling subset too. The other triviality is that the shortest cofinalling subset of any ordinal has to be a cofinal. Indeed, if it were not and thus had a shorter cofinalling subset then that were cofinalling subset of the original set too.
Which would contradict that we regarded the shortest cofinalling subset.
This means that we can call the shortest cofinalling subset of an α ordinal the cofinal "of " α. The cofinal of a cofinal is of course itself.
The cofinal of $\alpha + n$ is 1, of $\alpha + \omega$ and of ω_ω is ω and of ω_1 is ω_1.
Now to prove that all cofinals are cardinals we show the following:
For any ~ equivalence between two W , W' well-orderings, there is a V cofinalling subset of W that the ~ equivalence regarded only on V is a similarity.
First let's see why this implies that cofinals are cardinals.
Suppose α is a cofinal, so has no shorter cofinalling subset but were not a cardinal and so were equivalent to a β beginning of it.

Then our claimed V cofinalling subset of α were similar to a V' subset of β.
But α is similar to V so were also similar to V' that is shorter than β.
So now we show how to find the V cofinalling subset claimed above.
It should be those v elements of W to which the ~ equivalence assigned v'-s are such that to no $w \succ v$ is assigned a $w' \prec v'$. On V the equivalence is a similarity.
We must only show that V is a cofinalling subset of W. Suppose it weren't.
Let's pick a w_1 after V. By not being in V, there is $w_2 \succ w_1$, so that $w_2' \prec w_1'$.
By w_2 again not being in V there is a $w_3 \succ w_2$ so that $w_3' \prec w_2'$.
This w_3 is again not in V so we can pick a w_4. And so on, we would have a backwards infinite sequence in W' contradicting the Q minimality.
Back to the second more subtle problem, observe that ω is a limit cardinal and cofinal too but then ω_ω the next limit cardinal is obviously not cofinal.
So the question is whether much later we can find again cardinals that are like ω that is being a limit cardinal and being cofinal too?
These would be the limits of only as many cardinals as are themselves.
Which is trivial for ω because the earlier cardinals are just the finite naturals.
But for using infinite cardinals already as members this seems as a huge restriction!
And this hugeness is a plausibility, that is a priori decidable by the human mind.
Now comes the amazing coincidence that the existence of such cardinals is not derivable by the axioms of Set Theory. Finally some intuitive observations:
A W being equivalent to ω means that W is an other, non minimal or stretched well-ordering of the naturals. So ω_1 as well-ordering is very concrete!
It is the set of all possible material stretchings of the natural numbers.
They can not be combined into a maximal material stretching because their total as ordering is ω_1 and it is not equivalent to ω any more.
This set of the material stretchings then obeys the inside monotone unstretchability of any beginning to go all the way. A strange parallelity of the two stretchings!
These minimally well-ordered sets, the cardinals raise an other interesting question.
How can they emerge from the existence guaranteeing choice functions?
Obviously, the choices must be very very lucky to get such.
The fundamental coincidence is that the cofinalling subsets are also chance related.
Indeed, more generally, any subset can be regarded as the head or tail outcomes of coin flips at all members of a set.
So we must enter binary well-orderings and more generally Random Set Theory!
I believe that the "holy grail" the continuum problem lies behind such new theory.
To most mathematicians it will come as biggest surprise since problems are taboos.
But hiding even the easily understandable present plausibilities is worse than taboo.
Ordinals, cardinals, cofinals you will find on Wikipedia but the insane Formalists can turn even these beautiful ideas into a monstrosity of "regularity and singularity".
Poor Paul Cohen and Leonard Cohen didn't know yet that "Everybody knows that the Wiki is lying".

Appendix 7.: Well-orderings of the naturals

The abstract considerations may overshadow the amazing fact how particular well orderings come about by the different choice functions in the set of all A subsets.
So now I will show how the different choice functions in the case of the simplest infinite set, the naturals as our A set, generate the different well orderings of A.
The very start should be to see that already any reordering of A is actually a different choice and widening function.
For example, 5 , 3 , 2 , 1 , 4 , 6 , 7 , 8 , . . means that:

c (A) = a = 5 , c (A–{5}) = f{5} = 3 , c (A–{5,3}) = f{5,3} = 2 , . .

The real point is of course that the widenings can include material stretchings.
That is, restarting the infinity like the simplest 1 , 3 , 5 , . . 2 , 4 , 6 , . . early recognition of Galileo. The start and the f widening function for this is:

s = 1 , f {1} = 3 , f {1 , 3}= 5 , . . f {1 , 3 , . . } = 2 , f { 1 , 3 , . . , 2 } = 4 , . .

We automatically, or more exactly, a priori recognize that the essence is the new well ordering types obtainable.

So 1 , 2 , 3 , . . = ω while 1 , 3 , 5 , . . 2 , 4 , 6 , . . = ω,ω = ω+ω.

The point is that the "boring" reorderings must be part of the exact and thus specific realizations of these stretchings. Once we start to think only in types then these "boring" reordering variations of the actual elements tend to be forgotten.
But now I will show that there is a very visual representation of these particular stetchings. To make this even more surprising, I will start with a seemingly trivial alternative interpretation of any well ordering of 1 , 2 , 3 , . .

Namely, as a well ordering of the $\frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}$, . . = 1 infinite sum.

The above mentioned 1 , 3 , 5 , . . 2 , 4 , 6 , . . for example means that:

$$\frac{1}{2^1} + \frac{1}{2^3} + \frac{1}{2^5} + \;.\;.\; + \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \;.\;. \quad = 1.$$

Seems completely pointless yet.
Now let's regard the subsets of A as binary sequences.
For example, 0010111001…. defines the 3 , 5 , 6 , 7 , 10 , . . . subset.
Still nothing special, but let's place a decimal in front.
.0010111001…. is now a binary that defines a point in [0,1].

This goes the same way as for decimals but the combining from merely halvings is even simpler. Indeed, where we have a 1 value there we have a fraction with the corresponding power of 2 as denominator.
So for example: **.**0010111001…. = $\frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^6} + \frac{1}{2^7} + \frac{1}{2^{10}} + \;.\;.\;.$

The f widening function assigned a new outside natural to any set of naturals.
In binary sequence form this means to assign a new 1 in place of an earlier 0.
In the binary as infinite sum, it means to add a new 2 power fraction to our sum.
Now comes the surprise, that on the line this means a shift to the right.
So actually, our f is assigning to every P point of [0,1] a shift to the right.
This perfectly determines a transfinite shift sequence that moves 0 to 1.
This is the geometrical meaning behind the well orderings of $\frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3}$ . .
And so the possible stretchings of A correspond to possible movings of 0 to 1 in [0,1]. As usual, such nice new picture at once raises plenty of new questions too.

Our assigned right shift to a P in [0,1] was a yet missing 2 power fraction in P.
This ensured that P moved to the left of 1.
Allowing any 2 power shifts would stop this.

But this also suggests a "better" generalization by assigning to any point of [0,1] quite simply an arbitrary shift that only goes left of 1. Such assignment again would define an automatic transfinite shift sequence of 0 to 1.
But observe that every new "generalized" single move could be replaced by old move sequence using only yet missing fractions.
Indeed, all we have to do is apply moves but never surpass the single move.
Does this suggest that the well orderings of N can reach up to the continuum?
Not at all! Simply because here we are among the actual well orderings and not their types, that is their similarities.

Appendix 8.: Social Lies and Formalism

I clearly remember the following incident in 1967 when I was a first year math student in Budapest. I already fell in love with Set Theory in High School and hated the boring classical math subjects. I was reading Kalmar's text book and got entangled in his proof of the Well Ordering Theorem. It was several pages long.
A fellow student was Laszlo Suranyi the son of one of our professors Janos Suranyi.
I asked him about it but he just rambled something. With hindsight I don't know why I didn't turn to his father directly who was the nicest person.
Once Kalmar and Rozsa Peter sat on my sides in a special discussion and I didn't talk to them either. That kind of inhibited idiot was I at that age.
A year later I was at Rome waiting for my US visa reading Paul Cohen's Set Theory And The Continuum Hypothesis. That was the book I learned English from.
The proof of the Well Ordering Theorem was two lines in it!
I decided that I will tell him my objections of such a swift handling of something that deep. And I did! A year later in Stanford, where I was working in the Math Library waiting for him to return from England.
The details of what happened will be explained somewhere else.
We had a fight and the next day he came back to the Library and said that probably I was right but I have to struggle it alone and he is very different.
And indeed, he went on trying to solve the Riemann Hypothesis.
Instead of chewing on some remaining rotten flesh around the Axiom Of Choice.
But why did I delay the writing of this article for fifty years? I felt that some deeper details will resurface. And I was right, though they will not be presented yet.
Those, the random sets is the true reason why the Axiom Of Choice can not settle the Continuum Hypothesis.
But this "Seventh Appendix" will settle something much more important:
We live on a lying Planet!
Animals pretend for their survival but human lies are very different.
The existence of the individual soul could be the start of explaining this but it is too far from the concrete essence. While animals have individual souls just as we humans, they are in a much simpler trap than we. The wilderness kills their wonderful potentials. Their souls are not factors in their survival.
Only when they get in contact with us can their soul awake!
A tragedy! But ours is much deeper. Socialization awakens something in us too.
Thinking! Thinking about thinking can one make the simplest entrance into idealism.
Namely, that thinking can not be material. And presto, you are a Platonist.
What makes one push over this edge of the razor is varying.
For me it was Hegel's Lectures About The History Of Philosophy.
Already his starting sentence was making the razor's edge evident:
Gentlemen, we are starting to examine the history of philosophy, which is actually the history of idealism. Because materialism is not philosophy, it is bad science.
But this first sentence hides something deeper that Hegel never contemplated.

What is good science? It is the didactical awakening of the soul.
Kant contemplated very deeply how scientific understanding can come about.
He understood Newton's importance and tried to build it into an idealism.
But this was a fundamentally false fusion of science and philosophy.
And Hegel accepted the same oversimplified degradation of scientific understanding.
The hard fact is that Newton's physics, the first physics, is "bad science".
This sounds so provocative that it can lit a fire of at least asking "why".
The answer is simple. Beside validity there is truth! Well accepted by philosophers but never applied for Science. And the truth of Science is Understanding.
The a priory capacity of understanding science, is a jungle of concrete a priori plausibilities that have to be awakened one by one.
Each is a concrete proof of idealism and the soul's specifically human capacity.
So materialism is bad science only from the side of a philosopher while bad science can be science that is not even trying to be a philosophy.
The distinction can only come if one goes into science. Kant did, Hegel didn't.
But going into is not enough. One has to teach it. Transfer-trial awakens truth.
So the following joke is very serious:
The tutor explains something but the student still shakes his head. So the tutor starts from scratch, trying to fill in the missing details. After the student is still not happy, he exclaims: How the hell can you not see it when finally now I got it too.
But the discovering of scientific truth is an entirely different business!
It is purely a mercy of God! A G.F.G. a gift from God
Something given as either a maze of visions or a single obsession.
The "wide" geniuses receive it as the first. Like Newton, Gauss or Gödel, while the "deep" geniuses like Cantor, Einstein or Turing get the stuff that the seemingly smarter wide geniuses could not obtain. This duality is beyond us to understand.
But the point is that it is irrelevant too. Exactly due to the more important existence of didactics. This is the science of understanding valid for all humans.
But why is this a hidden and suppressed science?
That's where the social lies come in as the wider field that has to be unmasked first.
This more general duality is that not only the liars and patronizers are responsible but the ones who prefer to be lied to and to be patronized.
In our case, not only the Formalists are responsible who don't want to explain but the smart ass "learners" who want to become like their stupid masters.
In the even more concrete situation, the indigestibility of the Wikipedia science articles became not a trivial joke rather a sign of "excellence".
But speaking about "joke" is very appropriate because the side symptom of all these lies is being "funny" and accepting that things are never as they should be.
What these funny morons never contemplate in detail is how things should be!
My didactical awakening was very gradual.
From a Formalist moron only slowly became the militant anti Formalist.
Tutoring for need of money at uni started the process. I realized that my entrance into math in the special high school was not a didactical one at all.
I entered through Geometry that Plato wrote too above the entrance of the Academy.
But as I searched for help in my tutoring, I had to realize that a much better golden yellow brick road exists. It is the word problems.
Larichev's amazing collection was translated to all eastern block languages.
Later it was titled as Collection Of Algebraic Problems. And indeed, this expressed the unavoidable truth that we entered an age of algebraization. Of course, Geometry must be awakened too but it is a Garden not a Road. The third category, the Map is when we explain only its relations not the individual pieces.
These three must be used parallel but the essence is the establishing of Roads that give abilities and lead into a field.
Walking, talking, reading, writing, counting are the five basic abilities.

You either have it or not and it is immaterial how fast you achieved it.
We don't say how god someone walks except for models which is a special "walk".
The sixth Road, the multiplication of single digit numbers was questioned for a while when the so called complex model tried to make kids discover these multiplication results. It turned out to be a disaster.
Half of the "geniuses" could still not multiply at grade six.
So this mindless learning of the times-table is a must and it has subconscious roles.
The seventh Road is the word problems that incorporates indeed the algebraic rules, the calculations with fractions and decimals and the solving of simple equations.
But the true hidden plausibility that is awakened by all these is the use of variables.
Amazingly, still not crystallized by the time of Gauss!
There are three more Roads that should be introduced in Elementary Schools.
Grammatics, the use of quantors to rewrite everyday sentences that should not even be part of mathematics, Flow-charts and the Grid-system of the plane.
Instead of dwelling on the importance of these particular Roads, observe that the concept of obtaining abilities is a heuristic step in itself.
Just as I said earlier that there are no "good walkers", through all abilities we learn the most important thing that everybody is equal in the ability of understanding.
So school itself should be not a place of competitions rather a place where actually kids teach each other.

I learnt about Andre Toom who advocated the word problems in Sweden and South America without any breakthrough. Very sad story that made me even stronger.
In Sydney I tutored for years in the Newtown Library's Gallery.
After a while I only accepted kids who failed completely, to challenge myself and the poor parents who indeed experienced tear producing situations.
But to reform education itself seemed more and more impossible as I got deeper.
When I retired in Hungary, I went to the Pedagogical Library to see all the new text books published since I left. You wouldn't believe the monstrosities that I found.
A nice exception was Lajos Posa's math book for special high schools.
But a chapter started with statement that the Fundamental Theorem Of Arithmetic is beyond the level of the students.
I knew Posa from my high school time and also had a strange occurrence of his name in a letter from Erdös where he complained that such genius wasted his talent with teaching. My reply was even ruder than the earlier letter I wrote to him.
What a blind bat for Didactics. But I have to admit he admitted at some level the ideas that a laid upon him. Just as Cohen earlier in Stanford.

The concrete result of my reading Posa's view in his high school textbook became that I went again into the Unique Prime Factorization Theorem.
This became the most times rewritten article in my Didactical Gems section on my web site cognum.org. But for years now it is unchanged. A true didactical perfection.
The second most times rewritten article became this one about the well-ordering.
The third one is about the Twin Paradox.